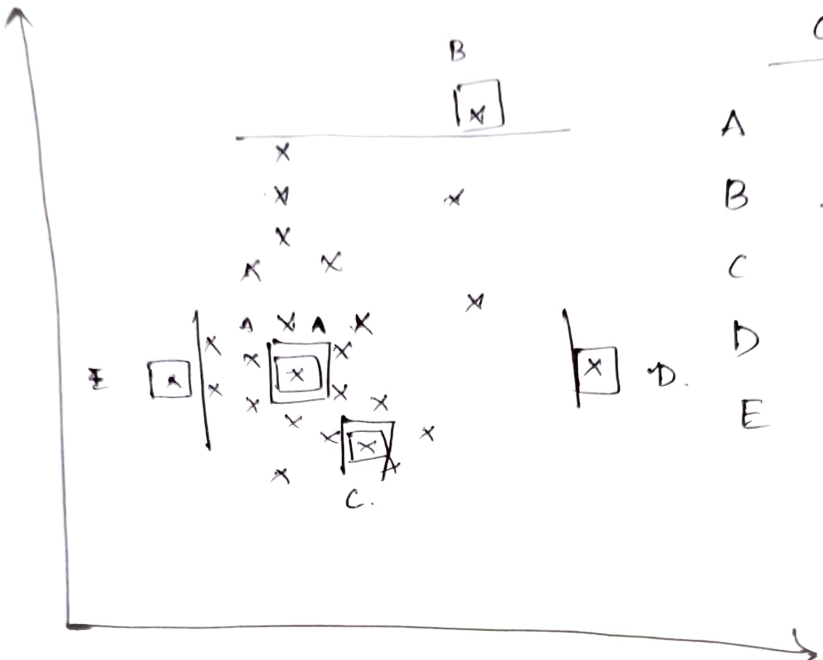
Isolation Forest

→ Unsupervised algorithm for outlier detection.



	cuts
A	4
B	1 ✓
C	3
D	1 ✓
E	1 ✓

The points with minimum cuts are considered as outlier points.

→ It ~~works~~ doesn't works for higher dimensionality of data.

Outlier Detection Methods	storage	speed	online	Approach
1) Z-score	Less	Low	✓	Mean & Standard devia
2) DBSCAN	More than Z-score	Medium	✗	Distance
3) KLE-LOF	More than DBSCAN	Medium	✗	Distance & Score
4) Isolation Forest	Less	Fast	✓	cuts

Expectation - Maximization

Tossing coin 10 times

(A) HHTTHHTHHHH

(B) HTTTTHHHTT

(A) HHHHHHHHHT

(A) HTHHHHTTHH

(B) HHHHHTTTTT

Events	coin A	coin B
1	7H, 3T	
2		4H, 6T
3	9H, 1T	
4	7H, 3T	
5		5H, 5T
Total	23H, 7T	9H, 11T

Bias of A w.r.to H? $\Big\} \rightarrow \theta_A = \frac{23}{23+7} = 0.76$

Bias of B w.r.to H $\Big\}$ $\theta_B = \frac{9}{9+11} = 0.45$

Tossing coin 10 times:

E_1 HHTTHTHHHH

E_2 HTTTTHHHTT

E_3 HHHHHHHHHT

E_4 HTHHHHTTHH

E_5 HHHHHTTTTT

Iter 1.

$\theta_A = 0.6$

Expectation $\theta_B = 0.5$

WKT, Events follow Binomial Distribution.

$P(E/\theta) = {}^{n}C_{x}\ \theta^{n}(1-\theta)^{n}$

Consider θ_A : $E_1 = \{7H, 3T\}$

$P(E_1/\theta_A) = {}^{10}C_{7}\ [0.6^{10}(1-0.6)^{7}]$

$= \frac{10!}{7!(10-7)!}\ [0.6^{10} \times 0.4^{7}]$

$= \cancel{0.00078}\ 0.00118.$

Consider θ_B :

$P(E_1/\theta_B) = {}^{10}C_{7}\ [0.5^{10}(1-0.5)^{7}]$

$= \frac{10!}{10!(10-7)!} \times 0.5^{10} \times 0.5^{7}.$

$= 0.00091$

$P(\theta_A) = \frac{0.00118}{0.00118 + 0.00091} = \frac{0.00118}{0.00209} = 0.56$

$P(\theta_B) = \frac{0.00091}{0.00118 + 0.00091} = \frac{0.00091}{0.00209} = 0.44$

consider E_2 $\{4H - 6T\}$

$$P(E_2 / \theta_A) = {}^{10}C_4 \left[0.6^{10} (1-0.6)^4 \right]$$

$$= \frac{10!}{4!(10-4)!} \times 0.6^{10} \times 0.4^4$$

$$= 210 \times 0.6^{10} \times 0.4^4$$

$$= 0.0325$$

$$P(E_2 | \theta_B) = {}^{10}C_4 \left[0.5^{10} \times 0.5^4 \right]$$

362 880

$$= 210 \times 0.5^{10} \times 0.5^4$$

$$= 0.0128.$$

$$P(\theta_A) = \frac{0.0325}{0.0325 + 0.0128} = \frac{0.0325}{0.0453} = 0.71$$

$$P(\theta_B) = \frac{0.0128}{0.0325 + 0.0128.} = \frac{0.0128}{0.0453} = 0.29$$

consider E_3 - $\{9H, 1T\}$

$$P(E_3 / \theta_A) = {}^{10}C_9 \left[0.6^{10} (1-0.6)^9 \right]$$

$$= \frac{10!}{9!(10-9)!} \times 0.6^{10} \times 0.4^9$$

$$= 10 \times 0.6^{10} \times 0.4^9$$

$$= 0.000015$$

$$P(E_3 / \theta_B) = {}^{10}C_9 \left[0.5^{10} (1-0.5)^9 \right]$$

$$= 10 \times 0.5^{10} \times 0.5^9$$

$$= 0.000019$$

$$P(\theta_A) = \frac{0.000015}{0.000015 + 0.000019} = \frac{0.000015}{0.000034} = 0.44$$

$$P(\theta_B) = \frac{0.000019}{0.000015 + 0.000019} = \frac{0.000019}{0.000034} = 0.56$$

consider $E_4 = \{7H, 3T\}$

$$P(E_4/\theta_A) = {}^{10}C_7 \left[0.6^{10} (1-0.6)^7\right]$$

$$= \frac{10!}{7!(10-7)!} \times 0.6^{10} \times 0.4^7$$

$$= 0.00118$$

$$P(E_4/\theta_B) = {}^{10}C_7 \left[0.5^{10} (1-0.5)^7\right]$$

$$= \frac{10!}{7!(10-7)!} \times 0.5^{10} \times 0.5^7$$

$$= 0.00091$$

$$P(\theta_A) = \frac{0.00118}{0.00118 + 0.00091} = \frac{0.00118}{0.00209} = 0.56$$

$$P(\theta_B) = \frac{0.00091}{0.00118 + 0.00091} = \frac{0.00091}{0.00209} = 0.44$$

consider. $E_5 = \{5H, 5T\}$.

$$P(E_5/\theta_A) = {}^{10}C_5 \left[0.6^{10} (1-0.6)^5\right]$$

$$= \frac{10!}{5!(10-5)!} \times 0.6^{10} \times 0.4^5$$

$$= 252 \times 0.6^{10} \times 0.4^5$$

$$= 0.01560$$

$$P(E_5/\theta_B) = \{ {}^{10}C_5 [0.5^{10} (1-0.5)^5]$$

$$= \frac{10!}{5!(10-5)!} \times 0.5^{10} \times 0.5^5$$

$$= 252 \times 0.5^{15}$$

$$= 0.0076$$

$$P(\theta_A) = \frac{0.0156}{0.0156 + 0.0076} = \frac{0.0156}{0.0232} = 0.67$$

$$P(\theta_B) = \frac{0.0076}{0.0156 + 0.0076} = \frac{0.0076}{0.0232} = 0.33$$

Events	coin A	coin B
E_1	$\theta_A = 0.56$ {7H, 3T} $0.56 \times 7H$, $0.56 \times 3T$ $= 3.92H$, $1.68T$	$\theta_B = 0.44$ {7H, 3T} $\rightarrow 0.44 \times 7H$, $0.44 \times 3T$ $= 3.08H$, $1.32T$
E_2	$\theta_A = 0.71$ {4H, 6T} $\rightarrow 0.71 \times 4H$, $0.71 \times 6T$ $= 2.84H$, $4.26T$	$\theta_B = 0.29$ {4H, 6T} $\rightarrow 0.29 \times 4$, 0.29×6 $= 1.16H$, $1.74T$
E_3	$\theta_A = 0.44$ {9H, 1T} $\rightarrow 0.44 \times 9H$, $0.44 \times 1T$ $= 3.96H$, $0.44T$	$\theta_B = 0.56$ {9H, 1T} $\rightarrow 0.56 \times 9H$, $0.56 \times 1T$ $= 5.04H$, $0.56T$
E_4	$\theta_A = 0.56$ {7H, 3T} $\rightarrow 0.56 \times 7H$, $0.56 \times 3T$ $= 3.92H$, $1.68T$	$\theta_B = 0.44$ {7H, 3T} $\rightarrow 0.44 \times 7H$, $0.44 \times 3T$ $= 3.08H$, $1.32T$
E_5	$\theta_A = 0.67$ {5H, 5T} $\rightarrow 0.67 \times 5H$, $0.67 \times 5T$ $= 3.35H$, $3.35T$	$\theta_B = 0.33$ {5H, 5T} $\rightarrow 0.33 \times 5H$, $0.33 \times 5T$ $= 1.65H$, $1.65T$
Total	$17.99H$, $11.41T$	$14.01H$, $6.59T$

$$Q_A = \frac{17.99}{11.99 + 11.41} = \frac{17.99}{29.4} = 0.61$$

$$Q_B = \frac{14.01}{14.01 + 6.59} = \frac{14.01}{20.6} = 0.68$$

Iteration - 2

With $Q_A = 0.61$, $Q_B = 0.68$

$$P(E/Q) = {}^{n}C_{x}\, Q^{n} (1-Q)^{n}$$

consider Q_A : $E_1 - \{7H, 3T\}$

$$P(E_1/Q_A) = {}^{10}C_{7} \left[0.61^{10} (1-0.61)^{7}\right]$$

$$= \frac{10!}{7!(10-7)!} \times 0.61^{10} \times 0.39^{7}$$

$$= 120 \times 0.61^{10} \times 0.39^{7}$$

$$= 0.00118$$

$$P(E_1/Q_B) = {}^{10}C_{7} \left[0.68^{10} (1-0.68)^{7}\right]$$

$$= \frac{10!}{7!(10-7)!} \times 0.68^{10} \times 0.32^{7}$$

$$= 120 \times 0.68^{10} \times 0.32^{7}$$

$$= 0.0008$$

$$P(Q_A) = \frac{0.0011}{0.0011 + 0.0008} = \frac{0.0011}{0.0019} = 0.57$$

$$P(Q_B) = \frac{0.0008}{0.0011 + 0.0008} = \frac{0.0008}{0.0019} = 0.43$$

consider $E_2 - \{4H, 6T\}$

$$P(E_2/Q_A) = {}^{10}C_4 \left[0.61^{10}(1-0.61)^4\right]$$

$$= \frac{10!}{4!(10-4)!} \times 0.61^{10} \times 0.39^4$$

$$= 210 \times 0.61^{10} \times 0.39^4$$

$$= 0.0346$$

$$P(E_2|Q_B) = {}^{10}C_4\left[0.68^{10}(1-0.68)^4\right]$$

$$= \frac{10!}{4!(10-4)!} \times 0.68^{10} \times 0.32^4$$

$$= 210 \times 0.68^{10} \times 0.324$$

$$= 0.0465$$

$$P(Q_A) = \frac{0.0346}{0.0346+0.0465} = \frac{0.0346}{0.0811} = 0.42$$

$$P(Q_B) = \frac{0.0465}{0.0346+0.0465} = \frac{0.0465}{0.0811} = 0.58$$

consider $E_3 - \{9H, 1T\}$

$$P(E_3/Q_A) = {}^{10}C_9\left[0.61^{10}(1-0.61)^9\right]$$

$$= \frac{10!}{9!(10-9)!} \times 0.61^{10} \times 0.39^9$$

$$= 10 \times 0.61^{10} \times 0.39^9$$

$$= 0.000041$$

$$P(E_3/Q_B) = {}^{10}C_9\left[0.68^{10}(1-0.68)^9\right]$$

$$= \frac{10!}{10!(10-9)!} \times 0.68^{10} \times 0.32^9$$

$$= 10 \times 0.68^{10} \times 0.329$$

$= 0.0000074$

$$P(Q_A) = \frac{0.000041}{0.000041 + 0.0000074} = \frac{0.000041}{0.000048} = 0.85$$

$$P(Q_B) = \frac{0.0000074}{0.000041 + 0.0000074} = \frac{0.0000074}{0.000048} = 0.15$$

consider $E_4 - \{7H, 3T\}$

$$P(E_4/Q_A) = {}^{10}C_7 \left[0.61^{10} (1-0.61)^7\right]$$

$$= \frac{10!}{7!(10-7)!} \times 0.61^{10} \times 0.39^7$$

$$= 120 \times 0.61^{10} \times 0.39^7$$

$$= 0.0011$$

$$P(E_4/Q_B) = {}^{10}C_7 \left[0.68^{10} (1-0.68)^7\right]$$

$$= 120 \times 0.68^{10} \times 0.32^7$$

$$= 0.0008.$$

$$P(Q_A) = \frac{0.0011}{0.0011 + 0.0008} = \frac{0.0011}{0.0019} = 0.57$$

$$P(Q_B) = \frac{0.0008}{0.0011 + 0.0008} = \frac{0.0008}{0.0019} = 0.43$$

Consider $E_5 - \{5H, 5T\}$

$$P(E_5/Q_A) = {}^{10}C_5\left[0.61^{10}(1-0.61)^5\right]$$

$$= \frac{10!}{5!(10-5)!} \times 0.61^{10} \times 0.39^5$$

$$= 252 \times 0.61^{10} \times 0.39^5$$

$$= 0.0162$$

$$P(E_5/Q_B) = {}^{10}C_5\left[0.68^{10}(1-0.68)^5\right]$$

$$= \frac{10!}{5!(10-5)!} \times 0.68^{10} \times 0.32^5$$

$$= 252 \times 0.68^{10} \times 0.32^5$$

$$= 0.0178$$

$$P(Q_A) = \frac{0.0162}{0.0162+0.0178} = \frac{0.0162}{0.034} = 0.47$$

$$P(Q_B) = \frac{0.0178}{0.0162+0.0178} = \frac{0.0178}{0.034} = 0.53$$

Events	coin A	coin B.
E_1	$Q_A = 0.57$ $\{7H, 3T\}$ $\rightarrow 0.57 \times 7H,\ 0.57 \times 3T$ $= \cancel{4.27H},\ \cancel{1.83T}$ $3.99H,\ 1.71T$	$Q_B = 0.43$ $\{7H, 3T\}$ $\rightarrow 0.43 \times 7H,\ 0.43 \times 3T$ $= \cancel{4.76H},\ \cancel{2.04T}$ $= 3.01H,\ 1.29T$
E_2	$Q_A = 0.42$ $\{4H, 6T\}$ $\rightarrow 0.42 \times 4H,\ 0.42 \times 6T$ $= 1.68H,\ 2.52T$	$Q_B = 0.58$ $\{4H, 6T\}$ $\rightarrow 0.58 \times 4H,\ 0.58 \times 6T$ $= 2.32H,\ 3.48T$
E_3	$Q_A = 0.85$ $\{9H, 1T\}$ $\rightarrow 0.85 \times 9H,\ 0.85 \times 1T$ $= 7.65H,\ 0.85T$	$Q_B = 0.15$ $\{9H, 1T\}$ $\rightarrow 0.15 \times 9H,\ 0.15 \times 1T$ $= 1.35H,\ 0.15T$
E_4	$Q_A = 0.57$ $\{7H, 3T\}$ $\rightarrow 0.57 \times 7H,\ 0.57 \times 3T$ $= 3.99H,\ 1.71T$	$Q_B = 0.43$ $\{7H, 3T\}$ $\rightarrow 0.43 \times 7H,\ 0.43 \times 3T$ $= 3.01H,\ 1.29T$
E_5	$Q_A = 0.47$ $\{5H, 5T\}$ $\rightarrow 0.47 \times 5H,\ 0.47 \times 5T$ $= 2.35H,\ 2.35H$	$Q_B = 0.53$ $\{5H, 5T\}$ $\rightarrow 0.53 \times 5H,\ 0.53 \times 5T$ $= 2.65H,\ 2.65T$
Total	$19.66H,\ 9.14T$	$12.34H,\ 8.86T$

$$Q_A = \frac{19.66}{19.66 + 9.14} = \frac{19.66}{28.8} = 0.68$$

$$Q_B = \frac{12.34}{12.34 + 8.86} = \frac{12.34}{21.2} = 0.58$$

$\theta_A = 0.68$, $\theta_B = 0.58$.

The θ_A & θ_B values of the iteration 2 are not similar to the previous iteration, Hence, the algorithm is not yet converged.